

LB

1131

J7

EXCH. 1102  
ARR. 9 218

UC-NRLF



\$B 263 025

The University of Chicago

# The Influence of Age and Experience On Correlations Concerned With Mental Tests

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL  
OF ARTS AND LITERATURE IN CANDIDACY FOR THE  
DEGREE OF DOCTOR OF PHILOSOPHY  
(DEPARTMENT OF PSYCHOLOGY)

BY

EDWARD SAFFORD JONES

Assistant Professor of Psychology, Oberlin College

A Private Edition

Distributed By

The University of Chicago Libraries

A Trade Edition is

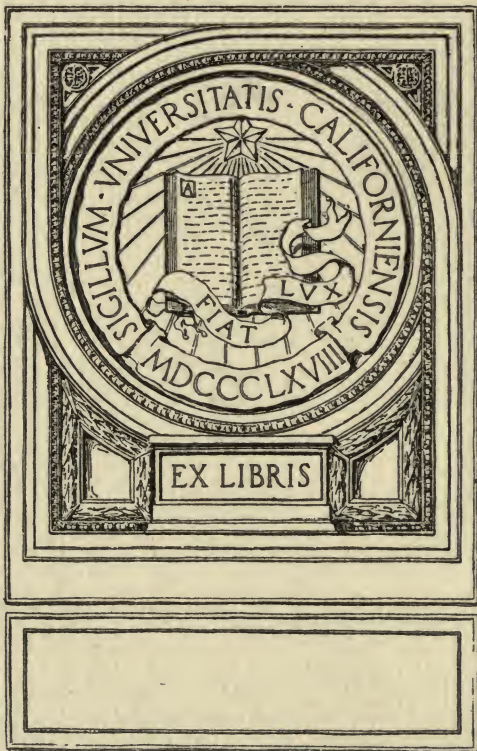
Published By

WARWICK & YORK, Inc.


Baltimore, Md.

1917

EXCHANGE







Digitized by the Internet Archive  
in 2007 with funding from  
Microsoft Corporation

**The Influence of Age and Experience  
On Correlations Concerned  
With Mental Tests**



The University of Chicago

---

The Influence of Age and Experience  
On Correlations Concerned  
With Mental Tests

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL  
OF ARTS AND LITERATURE IN CANDIDACY FOR THE  
DEGREE OF DOCTOR OF PHILOSOPHY  
(DEPARTMENT OF PSYCHOLOGY)

---

BY

EDWARD SAFFORD JONES

Assistant Professor of Psychology, Oberlin College

---

A Private Edition  
Distributed By  
The University of Chicago Libraries

A Trade Edition is  
Published By  
WARWICK & YORK, INC.  
Baltimore, Md.

1917

*[Edward Safford Jones]*

24 JLI8

LB1131  
J7

Copyright, 1917, by  
WARWICK & YORK, INC.

WARWICK & YORK, INC.



## CONTENTS

	Page
1. Introduction .....	3
2. Historical .....	7
3. Administration of the Tests.....	25
4. Results: The Influence of Age and Experience on—	
(1) Correlations of Samples of the Same Test on Different Years.....	41
(2) Relationship of Total Test Average and the Different Test Measure- ments .....	51
(3) Change in Inter-Test Correlations....	57
(4) Relationship of School-Grade—com- pleted to Test Measurements.....	67
(5) Influence of Education on Inter- Test Correlations .....	74
(6) Meaning of Individual Tests—Sum- mary .....	76
5. Conclusions .....	81
Appendix: List of References, with Table for Determining Probable Errors.....	87



## ACKNOWLEDGMENTS.

Valuable suggestions as to methods of handling the data of this research were received from Messrs. Harold D. Rugg and Beardsly Ruml, of the University of Chicago. In much of the routine connected with the computing of correlations, as well as in the writing of the text, my wife, Mrs. Frances Jeffery Jones, was of inestimable aid.

To Mrs. H. T. Woolley, with whose permission this study was carried on, credit for the collection of data is due. Without the steadfast planning and the inspirational guidance which came from her, no such body of data could have been gathered. Amid almost insuperable difficulties she has insisted on keeping the testing up to a high degree of standardization, allowing only skilled and interested workers to do the testing, and cheerfully but doggedly insisting each year on as full a quota as possible of "re-tests" of our industrial group. Few will realize the obstacles she has faced.



# THE INFLUENCE OF AGE AND EXPERIENCE ON THE CORRELATIONS CONCERNED WITH MENTAL TESTS.

## SECTION I.

### INTRODUCTORY.

In this research\* we have covered the data of four years of testing in the Vocational Bureau of Cincinnati. The results of two of these years are included in the Woolley and Fischer monograph, but we have limited ourselves in a number of ways. In the first place, our problem is more specific. We are interested in only one phase of the whole experiment, namely, the correlations between purely mental measurements, and their variations from year to year. In the second place, we have selected the records of boys only, and of those boys whose records were complete and continuous through the four years of testing. Thirdly, we have restricted ourselves to those tests which are called "mental," paying no attention to such tests as seem to depend more on physical ability. Finally, we have paid lit-

---

\* The present study is a by-product of the work of the Cincinnati Vocational Bureau, under the direction of Mrs. Helen Thompson Woolley. For an extended report of the problems and methods of that larger social experiment, one is referred to Monograph No. 77 of the *Psychological Review Monographs*, and to "A New Scale of Mental and Physical Measurements for Adolescents, and Some of Its Uses," in the *Journal of Educational Psychology* for November, 1915. These reports include also the age and grade norms in the different mental tests as administered by the Cincinnati Vocational Bureau.

the attention to tests which have not been carried on from year to year in much the same form. In other words, our emphasis was not on the variety of tests and a study of individual tests, but rather on the change in the various capacities measured by certain tests from year to year as indicated by their correlations.

More specifically, we are interested in answering the following questions:

1. On the basis of our standard psychological tests, do persons tend to become more alike from year to year? If they do tend to become more alike in their mental capacities, the correlations between these tests will be smaller from year to year; whereas if adolescent boys become separated more sharply into the good, mediocre or poor intellectually, the correlations between the tests will rise. (See note on p. 14.)

2. In the case of each of these standard tests, is the first, or a subsequent, testing of an individual the more reliable index of the tested capacity?

3. What tests are best correlated with the amount of school training received by the subjects of the tests? How does the relationship between school-grade-completed and the various tests change from year to year under the conditions of the Cincinnati testing?

4. What tests are the most reliable measures of intelligence over long intervals of time, as con-



trasted with tests giving highly consistent correlations for short periods? In what sorts of measurements do the persons tested hold closest to type from year to year?

5. What bearing have these results upon the presence of a common intelligence factor?

We have restricted ourselves entirely to the correlation method of procedure in getting our conclusions, because it seems to be the most adequate means for studying such relationships. It has frequently been said that the method of correlation can prove anything, and we are aware of the pitfalls of a wholesale, uncritical use of the method. The correlation index between Mental Test A, and Function X may be shown to be anywhere from 0 to plus .50, depending upon a great variety of factors not as a rule considered in the statement. Some of these factors are the age of persons tested, the sex, the educational status of the group considered, the exact method of giving the test (i. e., Mental Test A), familiarity with the test through practice, and finally, perhaps most important of all, the homogeneity of the group. Those who have worked to any extent in the field of correlational psychology realize the absurdity of such a statement as that "the Opposites Test correlates as high as .60 with English ability, or salesmanship, or what not." We shall have to be more explicit as to the conditions affecting such correlations, adding, for exam-

ple, that the group tested was one of college Freshmen in a Middle Western State Institution who had had one previous set of tests a month before, that the Opposites list was comprised of hard words and given according to the X method, and that English ability was based on such and such an attainment.

We do not anticipate that, in the following pages of results, there will be many individual correlations which will have concrete value to those anxious to apply the tests for vocational purposes. We do believe, however, that the data which we are offering are capable of adding to our present information regarding the very complex set of factors which make up the average set of correlation indices.



## SECTION II.

### HISTORICAL.

Researches concerned with the correlation method of procedure in treating of the results of mental tests have been quite fully described in many places, and are now so numerous that no attempt will be made to survey the literature entirely. Readers are advised to read such articles as those by Whitely ('11)\*, Brown ('11), Hart and Spearman ('12), and Thorndike ('14), to become acquainted with the origin of "correlational psychology," and also to come in contact with practically all the statistical technique now used by psychologists in discussing mental tests. Before taking up to some extent those special researches which bear quite directly on our problems, it might be well to summarize briefly the general results which have been pretty well agreed upon by all who have worked extensively in this field.

1. Correlations between all sorts of mental measurements of a desirable nature and with all kinds of groups of subjects, tend to turn out positively. In the case of most experiments, those tests which appear to measure the higher mental capacities, as contrasted with tests of abilities on a purely sensory level, correlate with each other highly.

2. The degree of correlation between mental tests varies markedly according to the group of subjects tested (even where the same tests have been

---

\* This refers to a publication in the year 1911.

used and by the same experimenter), showing:

(a) the dependence of the degree of correlation on the heterogeneity of the group, and

(b) on the tests' meanings to subjects of different age and educational status.

3. For the most part, those tests which appear to measure closely related abilities have given high correlations when tried out together, but numerous cases to the contrary have been cited. For example, Thorndike ('14), and Winch ('09), and Wyatt ('14), have concluded that memorizing abilities are highly correlated, more on account of like content than because of the mental process. The memorizing of two entirely different materials may show no positive correlation.

4. A fourth result might be added to the outcome of mental test researches. There has been more and more evidence to indicate, as Spearman holds, that a correlation, practically perfect, exists between columns of correlation indices of mental tests. This has been taken to prove that tests can be arranged in the order of their ability to measure a hypothetical common factor of intelligence. Some writers, however, notably Simpson and Brown, have found it difficult to subscribe to such an arrangement of their tests.

In searching for an historical background for our study, we find the data very meagre. However, many important researches have touched on one or

another phase of it. In enumerating them, it would seem best to classify them under the following divisions:

1. Those bearing on the influence on the correlations of a difference in the subjects tested.
2. On the influence of practice on correlations.
3. On the effect of time intervals on correlations of samples of the same tested ability.

(1) *Those researches discussing the influence of age, intellectual status, or homogeneity of the group on the correlations between mental tests.*

Mr. C. Scott ('13) believes that one would find in the average normal school that estimates of intelligence and ranking in tests correspond most highly at the time of the first year of normal school study. He believes this would be true because it is the period just before specialization has set in. On such an hypothesis, one would expect a higher correlation between tests of intelligence among persons of fourteen years of age, especially when they have just come out of the public schools, than among the same persons tested after work in the industries. Specialization and the levelling influence of industrial life might well operate to make many persons fair performers in certain mental tasks and inefficient in others, thus lowering the correlation indices.

On the other hand, it might be argued that school work was the more levelling type of intellectual

performance, and that when once away from its routine the brighter persons would go ahead much more rapidly than before, and the duller ones, because of the lower grade of their jobs, would stay practically where they were.

Burt ('09) seems to have been the first to apply the same set of tests to more than one group of subjects. He gave his set of twelve tests first to a group of thirty elementary school boys between the ages of eleven and thirteen; he repeated these tests on a group of thirteen preparatory school boys of about the same age. The distinction between the two groups of subjects was not a difference in age, but a distinction on the basis of social standing, determined by the wealth of the parents. As would be expected, there was more selection on the part of the preparatory school boys, and consequently a greater homogeneity of intellectual abilities. This was shown by the lower correlations between the tests in the study of the preparatory school group. Out of 77 compared correlations between tests taken by elementary, as well as preparatory, school pupils, it was found that 47 of the elementary school correlations were higher than those of the preparatory school, while in one case both were the same. But the correlations connected with the preparatory subjects are somewhat unstable, because of the small number of subjects tested.

Of more interest than the comparison between these two sets of intercorrelations, is the compari-

son between the correlations obtained by Burt on his elementary pupils and the correlations of Calfee ('13), who worked with four of Burt's tests on a group of college freshmen in the University of Texas. In every case, the correlations found by Calfee, on testing the freshmen, were decidedly below Burt's correlations with the same tests.

Definitely then, the selection which is present in the case of college freshmen, where there is marked specialization and also elimination of the unfit, lowers the correlations.

The more recent research of Bell ('16), in which there was scarcely a correlation index over .30, is another instance of the homogeneity of the college group.

Brown ('10) reports a study of a series of mental tests tried out on different groups of students, the lowest two composed of girls and boys of eleven and twelve years of age, and the upper two of university students, men and women. In practically every case, the coefficients of variation, or measures of the dispersion of the test records, were smaller with the older groups of subjects, suggesting the greater homogeneity of the older groups. Those correlations between tests, which were compared from one year to the next, showed also that in the older, more selected set of subjects the correlation between tests was lower, indicating greater homogeneity. Simpson ('10), working with two groups of subjects composed one of college graduates and



the other of derelicts, on combining the two groups, found the correlations higher than when they were considered separately.

The study of Bonser ('10) on the *reasoning ability of children* is the first large attempt to study the variation in correlations of mental tests in different age and grade groups. He tested 757 children from the fourth, fifth, and sixth public school grades. The tests were fairly complex, dealing largely with the language function, and were classified under the four captions of *mathematical judgment*, *controlled association*, *selective judgment* (including the Opposites Test, the only test which compared directly with any of ours), and the ability referred to as *literary interpretation*. All but the last type of mental tests showed reliable correlations with age and school grade.

One fact of great interest to us in Bonser's study is the degree of correlation between the tests and the school standing in the different grades. It turned out that the higher grades gave better correlations between school grade and the tests, but he affirms that this is due to the greater heterogeneity of the higher grades, as there was better distribution of mental ability and wider age variation in the upper grades than in the lower. This last factor explains, perhaps best of all, the divergence of these results from the results noted above, namely, that lower test intercorrelations occurred in the higher school grades.

Bonser also determined the correlations between tests in different age groups, particularly the three groups of "lowest 25 per cent. in age," "medium 50 per cent.", and "highest 25 per cent." He found results this time that were comparable to the above-cited findings, namely, that the correlations in the case of the younger subjects were somewhat higher than the corresponding correlations with older subjects.

Abelson ('11) worked with a number of tests on sub-normal or backward children, and came to this conclusion among others, i. e., that it is wrong to infer the value of a test in one group of subjects on the basis of the correlation connected with it in other groups. A test might have a fair degree of validity as a measure of intellectual capacity on one level of ability or with one age, and yet be mere routine, not even challenging real intellectual insight, on another level of age or mental ability. Such a statement, also mentioned by Brown ('11), would seem to challenge the validity of many conclusions previously mentioned. Naturally, it might be said, the correlations in connection with college students would be lower than those in the case of boys, because the tests which were hard mental performances for the boys, thus ranking most reliable in sorting out the good from the bad, would be uninteresting and intellectually not stimulating for college students.

Pyle brings out the same point in his study of the mental records of over 200 students, classified according to their age from eight to eighteen. The tests, which were valuable in differentiating age groups among the younger subjects, were not so effective in this respect with subjects of more advanced age.

These conclusions as to the influence of age on correlation seem to be somewhat uncertain. On the whole, it would appear that, when we go up the scale in age, especially when considering those tested in school, selection and mental homogeneity are evidenced by a drop in the correlations between mental tests\*.

---

\* Objections might be raised to the argument that low correlation tends to be an accompaniment of homogeneity among subjects. In the ordinary product-moment formula,

$$r = \frac{\Sigma(xy)}{n \sigma_x \sigma_y}, \text{ homogeneity reduces the size of the devia-}$$

tions  $x$  and  $y$  and, consequently, the numerators of the fraction. Now it is argued that homogeneity also insures a decrease in the size of the standard deviations, which make up the denominator of the fraction. The result would be no guaranteed change in the correlation.

This argument would be convincing if there were a straight line regression and only slight deviations from the means in each array. But this is never the case with mental tests. If one artificially chops off the highest and dullest of a large group of persons, it eliminates those individuals who have extreme deviations in both functions considered. This strikes off with a blow those affecting the correlation most positively. This artificial division is, however, not likely to interfere with those individuals extremely good or poor in one function, and mediocre in the other. All such cases operate to keep the S. D's as



The advantage of our results, in regard to this question, is that we dealt with exactly the same subjects each year, and not with groups of subjects, differing as to degree of selection operative among them and the amount of school work completed. Moreover, the age limits, from fourteen to seventeen years, are not wide enough to raise the objection that the tests have radically different meanings at one period of testing as compared with another.

Our data, however, have one serious disadvantage in attempting to decide the question of the influence of age and experience on the intercorrelations between tests, i. e., the effect of practice is involved in all tests beyond the first series given at the age of fourteen. The question arises, can we estimate the degree of practice and eliminate its effect in our work? This brings us to the next section.

(2) *Those researches discussing the effect of former experience and practice on the correlations between mental tests.*

The problem does not seem to have been directly attacked by many experimenters, but a number of researches have taken up the closely related ques-

---

high as before, whereas the  $x$   $y$ 's are negligible; i. e., the denominator is not affected as much as the numerator.

We have tried this artificial selection in the case of the Cincinnati subjects. When we disregard a few markedly inferior or superior individuals, whether in scholarship or in average mental test standing, the result is a smaller correlation between tests.

tion as to whether those who are good in mental tests are apt to improve as much, or more, than those who are poor. Obviously, if the poor improved with practice much more than the good, the group would tend to even up in ability, and the correlation between practiced tests would be reduced. If, on the contrary, those who are superior in initial test performances improve the more on account of practice, it would seem that correlation between practiced tests (assuming the validity of a general intelligence factor) would increase.

Binet takes the first of these positions, particularly on the basis of an experiment carried out with the five brightest and six dullest of a group of thirty pupils. With the cancellation test, tried out at four different times, he found that there was greater improvement with practice in the case of the dull pupils than in the case of the bright. This was much more marked in regard to accuracy than to speed. He concluded that the differentiation between the good and poor, in the case of mental tests, diminishes and tends to disappear entirely with continued testing of the same function. If this was true with all tests, as suggested by Binet, with continued practice the correlation between tests would approach zero.

The next research to take up this problem is that of Kruger and Spearman ('06), which works over the data on continued adding for a period of two hours, obtained by Oehrn in 1889. The standing

of each of the subjects for efficiency during the entire performance was compared with their standings during each of the fifteen-minute periods of the tests. In harmony with the doctrine of the common intelligence factor advanced by Spearman in 1904, their conclusion was to the effect that practice tends to increase the amount of divergence between the good and the poor, as evidenced by the rise in correlations between total efficiency and special efficiency on account of practice.

Burt ('09) is the next author to bring up the question of the influence of practice on correlation. He found that when from two to four trials of his tests were made on elementary grade students, the final trials correlated less closely with imputed intelligence than did the early trials. In all but one of the twelve types of measurements, he finds a lower correlation between test measurement and imputed intelligence on the second trial of the test than on the first trial. His conclusion was similar to Binet's main conclusion that continued practice with a test reduced its correlation with intelligence. If one will take note, however, of those correlations which concern tests administered three times, it appears that, although the third samples of ability correlate with intelligence more poorly than the first, yet the third samples are somewhat better than the records taken the second time. We want to emphasize, then, contrary to Burt's statement, that his data indicate a drop in the intercorrelations between

tests from the first to the second year and a rise from then on. This suggests that perhaps, in attacking a test for the first time, there is the factor of adaptation, common to all tests, which raises intercorrelations between tests on the first year higher than would otherwise be the case; and after the initial drop in intercorrelations, due to this factor common to all tests given for the first time, practice does operate to raise the intercorrelations from one time to the next.

Abelson ('11) tested subnormal children and concluded that, on the average, the intercorrelations between tests do not fall, but if anything, rise when the same series is repeated. The average intercorrelation between tests, given on a number of groups (boys and girls), turned out to be .32 for the initial trials, .36, .37, .40 in the case of the second, third and fourth trials, respectively.

Whitely ('11) reviews the work of many who have studied the effect of practice, and on the basis of these and her own results, derived from the records of only nine adult subjects, she concludes that individuals of low standing can and do improve more than those of high mental ability. Her tests were the discrimination of weights, cancelling A's, sorting, and the pencil maze. Her foot-rule correlation index, between position at the start and gross gain in the case of each of her tests administered twenty different times, was in the neighborhood of .50. Thorndike objects to these conclu-

tions on the basis of Whitely's other findings, and the data of Kirby, Starch, Wells, Thorndike and others. In none of these cases are the correlations designated in detail, but in every case there is indication of a heterogeneity of mental ability greater after a session of practice than in the initial trials. Thorndike ('14) concludes "the results are rather startling. Equalizing practice seems to increase differences. The superior man seems to have got his present superiority by his own nature rather than by superior advantages of the past, since during a period of equal advantage for all he increases his lead."

These conclusions gain slight verification again in a recently reported study of Thorndike (*Am. Jr. Psy.* 1916). A recent monograph by Wallin ('16) gives data from repeating form board tests on several hundred children, in which he found that the average pupils improved appreciably more than the duller or the bright, while the dull improved slightly more than did the bright.

Brown ('13) mentions an experiment by Winch concerning the effect of practice on correlation between a simple motor test of cancelling all letters, and a complex motor test, cancelling a,n,o,s. The correlations in the case of boys on the six successive days were .29, .44, .59, .48, .50, .47, suggesting an initial increase from the first to the second trial, but from then on no appreciable change.



In contrast to the rather dogmatic conclusions advanced by Thorndike to the effect that the better subjects do seem to improve in mental tests more than the poor, are the results of the recent researches by Wells ('15) and Chapman ('15). Wells found that those who gained most in adding over a period of thirty days did not gain the most in the Cancellation Test during the same period of practice. There was a high negative correlation between improvement in one case and improvement in the other case. Chapman worked with six tests on 22 college men, repeating each test ten times, and found no reliable correlations between improvements in the different tests, except in the two Cancellation Tests which were quite similar. In his correlations between initial standing and improvement, there were only negative or small positive  $r$ 's, except in the case of the adding and multiplying tests.

The last important contribution to the general question of the influence of practice on correlation is that made by Hollingworth in 1913. He tried the same six tests on each of thirteen subjects for 205 times, and found the correlations between the different tests at each of the points along the curve of learning. He found a tendency for correlations to be markedly higher towards the end of the practice curve than initially. He took the correlations of each test with each of the others at the following stages: (1) at the first trial, (2) the average

records of the first five trials, (3) the average of the twentieth to the twenty-fifth records, (4) the average of the 75th to 80th records, (5) the average of the 200th to 205th records inclusive. At each of these points the average intercorrelation of the tests used (Adding, Opposites, Color-naming, Discrimination, Co-ordination, and Tapping) was .065, .280, .320, .390, and .490 respectively. Unfortunately the comparison between the first and the second figures given above is not a valid one, as the measurements correlated were not comparable. In the first case, the measurements which were correlated were those of single trials (the first made), whereas in the following index averages five measurements were used for each correlation rather than one. Clearly, the correlations between such average records would be higher than those between single attempts. The rise in the correlation from the second figure on, is based on comparable data, but it hardly concerns our problem directly, as a practice of twenty-five consecutive times is not at all similar to two or three previous trials of a test. Also, as Hollingworth himself has pointed out, tests which have been gone through a large number of times would be bound to change radically because of constant habituation. (Compare an attack on the Opposites Test the fifth time with a similar attack when the test has been taken eighty times by the same subject).

The results bearing upon our question of the in-

fluence of practice on correlations between mental tests seem to be conflicting. On the basis of those experiments which deal with subjects similar to ours, and with tests comparable in nature, we feel justified in concluding that in the intercorrelations between mental tests there will be little change due to the succeeding practice effects of the Cincinnati tests. When tests are given a full year apart, it is not likely that any real practice effect will last over from one year to the next. In case there is a change due to practice, the major evidence seems to be in favor of a slight increase in the average intercorrelation between tests.

(3) *Those researches discussing the experimental evidence dealing with the amount of stability of intelligence measures, tried out at two different times.*

To make a distinction between a "reliability index" and a "stability index," the former refers to the correlation between two samples of the same test when these are taken soon after each other, and the latter refers to correlations between two samples separated by long periods of time.

The experiments on mental tests have frequently taken cognizance of the reliability of test measurements, but few have taken up the question as to how faithfully a test measurement will stand for a particular capacity over a long period of time.

Burt, Brown, Abelson, Simpson and others have worked with reliability indices between different



samples of the same test. A detailed study of these does not concern us, as the reliabilities of various tests were widely different, and none of these tests given were directly comparable with any in our series. Suffice it to say that as a rule those tests correlating most highly with the other mental tests or with imputed intelligence had the higher reliability indices.

Kruger and Spearman ('06) mention that in their tests reliability correlations were almost as high when the capacities were measured by different people a week apart as when measured by the same person twice in the same day.

Burt tried a series of tests on boys eighteen months after the initial series of tests, during a rapid growing period (thirteenth to fifteenth years), when interests and amount of knowledge absorbed varied greatly, and he found no corresponding variations in ability as measured by the tests. The capacities measured constituted relatively permanent endowments.

Starch ('13), in working on school grades, in the case of grammar school children found high consistency correlations (above .80), especially when grades were considered over a number of years in the grammar school. Kelley ('15) emphasized the long persistence of general and special abilities in the case of children in high school. Superiority in the grammar grades in a special subject such as mathematics was shown by the method of partial

correlations to have a definite counterpart in high school mathematics, the factor of general school ability remaining constant.

Wells ('15), working on Adding, Cancelling and Tapping Tests for thirty days and then dropping them until eight months later, found that in general those who gained the most through practice lost the most from disuse; but on the whole subjects held to their relative positions faithfully.

Mrs. Woolley (15), in her article already referred to, is the first to give a statement of the amount of correlation for the average mental test records on two consecutive years, given a year apart. For all children tested on the two years, there was a correlation of .71 with a probable error of .034. The physical tests gave a correlation somewhat lower, .64. The relationship between physical and mental tests on the first two consecutive years is also of interest in its bearing upon the effect of age and experience on tests in general. The correlation between the mental and physical series at fourteen was .21, whereas at fifteen years, it was .33.

In summary, it would appear that the ability of individual test measurements to determine a stable capacity over a period of time has not been carefully studied anywhere. General gross estimates of ability, such as school grades or averages in mental tests, seem to test capacities which are quite stable over long time periods.

### SECTION III.

#### ADMINISTRATION OF THE TESTS.

To come back to our own problem, it will clearly be seen that differences in methods of giving the tests, differences in the subjects tested, and in what is meant by practice and experience, are so various that there is need for us to define more precisely the exact nature of our age and experience factors. We are interested in the strictly adolescent period, between 14 and 18 years of age. It is usually assumed that the greatest changes in life—both physical and mental—appear during this period, so that if there be a noticeable change in the mental test relationships within short time-intervals, we ought to have basis for a generalization regarding it.

But we are not at all concerned with the factor of age isolated from that of general experience in the world of affairs. Obviously, to give everyone the same experience would be impossible. Hence it is necessary to consider the effect of age and experience as a combined factor. Furthermore, our conception of "experience" must include not only the vast sum of sensory-motor reactions peculiar to the Cincinnati industrial environment, but also previous familiarity with the test in question. To outline the factors concerned in our study, we might state that we are interested in knowing the influence on the correlations concerned with mental tests exerted (1), by an advance in years during the adolescent period of the average working boy of

Cincinnati; (2), by the series of chance experiences undergone by such a group during four years of contact with (a), widely different industrial occupations, and (b), different social conditions and miscellaneous educational influences affecting each boy to a different degree, and (c), the experience of coming to the work-certificate office, and under controlled conditions there going through a series of mental tests for four consecutive years (each year's results with the exception of the first year, influenced by one or more previous sets of similar tests).

Notwithstanding the vague and unverifiable character of our factor of experience, we feel that in the long run some general validity will be found in the results. The effect of age and experience on typical Cincinnati working boys will not be markedly different from the influence of the same factors elsewhere.

*The Subjects.\**

The subjects whose records are included in this study were 203 practically unselected boys, who started to work in the industries of Cincinnati at the age of fourteen. The work certificate office was equipped by Mrs. Woolley in such a way that each individual who came into the office to get a work permit at the age of fourteen was also given a series of mental and physical tests. This was re-

---

\* The complete records of every subject are on file at the Vocational Bureau of Cincinnati.



ardless of his education (except that no boys who had not completed the fifth grade were allowed to have certificates), regardless of nationality or race (except that no negroes were included), and regardless of his desire, or the desire of his parents, to have the tests given. There were 423 of these working boys who were tested in this way, and questioned as to their social, educational and industrial background. As many of them as possible were brought back each succeeding year to be re-tested and requestioned especially as to their industrial experiences. Surely the final comparison between physical, mental and other factors with their industrial and commercial progress will be of great value. Up to the time of the main work upon this research, the data as to industrial progress of these subjects had been gathered only for the first two years. The correlation up to the end of these two years between mental test proficiency on the one hand and the average of wage earnings and permanency of occupation on the other hand, was .07 and .11 respectively (by the Spearman foot-rule formula, taking 100 cases at random).

Two conditions were noticeable in bringing about this lack of correlation between measured intelligence and industrial fitness. In the first place many influential relatives helped poorly equipped individuals into the better paid jobs—jobs not so difficult, to be sure, as to demand much intelligence.

There was also a frequent tendency for the better class of boys to take up the poorly paid though substantial jobs to begin with, in spite of longer periods of apprenticeship and lower rate of wages. Both of these causes of low correlation will tend to be eliminated in time; but for the purpose of this thesis no note need be taken of industrial facts of whatever kind. It is hard to estimate accurately the degree of selection which was present in cutting down the original number, 423 unselected boys, to the 203 whose records are included in this research. But we feel that practically all this selection was of an accidental character, at least of a kind not connected with intellectual capacities. There were a few, not over 5 per cent., who did refuse to come back for one or another of the yearly test-series, and our 203 were composed only of those who had been tested on each of the four consecutive years. There was a larger per cent. of cases which were not included because of the incompleteness of test records, due to a number of factors such as poor stop-watches, or the insufficiency of experimenters in the office at rush periods; and only those were included who had taken every year at least two-thirds of the mental tests with which we were interested. Finally, the most important source of elimination, including at least 50 per cent. of all the rejected cases, was the fact that subjects moved out of town or left no trace of their moving to other parts of the city.

Just how much this selection is one connected with intellectual standing it is hard to determine. We must grant that those who refused at one time or another to come back for a re-testing were usually below the average in mental ability (as determined by previous tests). Also it is true that those whose addresses were lost on account of frequent moving were probably below the mental average of the total larger group, but this tendency is certainly not significant. An objective measurement of the amount of similarity between our 203 cases and the total 423 was made by comparing the average test records of the 203 with the original total group in the first year. The average for the 423, obtained by the Woolley method of average percentile rating (see page 37), was 54:70 per cent.; whereas in the case of the 203, the average of the first year turned out to be 55.97 per cent., with a standard deviation of 15.60 per cent. This shows that there is but a slight selection, so far as the intellectual capacity of our subjects is contrasted with the total group. All were fourteen years of age at the start, and were tested within two months of each subsequent year period for four years.

*The Tests and Methods Used.*

There were a large number of tests used throughout the four years of testing, but we decided for the purposes of this research to pay attention only

to those which might be designated as "mental," as contrasted with the "physical" series of tests. Mrs. Woolley ('15) has taken up a discussion on the separation of the tests along this line, and has also given a complete account of all the tests used, including scores and norms for each of the various tests. (See monograph No. 77, *Psy. Rev. Mono.*, and article in *Jr. Ed. Psy.*, Nov., 1915). She has laid great stress throughout on the importance of keeping strictly to specific directions in giving the tests. Those who are interested in administering any of the tests mentioned in this research we refer to the larger work of Woolley and Fischer (as above) for complete instructions. Not only have we restricted ourselves to the "mental" series of tests as distinct from the physical, but we have paid particular attention only to those tests which were repeated from one year to another. We have not dealt with the results of form-board tests, for example, because they did not correlate well with each other from year to year, and because, in many cases it seemed to the author that success in them depended largely on luck. Moreover, not until the third and fourth years did we get form-board tests which seemed to compare closely with each other. We were interested in no test which did not have its counterpart in other years. Otherwise our results would not be comparable from one year to another. It should be added, however, that in



every case slight changes have been introduced to forestall specific practice effects carrying over from one year to the next. (See below).

The following tests were chosen for specific study and measurements from these were used entirely: The Cancellation Test, a Substitution Test, Immediate Rote Memory for Numbers, and Cincinnati form of the Sentence Completion Test, and the Opposites Test. Miscellaneous correlations with a Cause and Effect Paired Associates Test (given in the third year in the place of the Opposites Test), and also with a Mutilated Text Completion test in the fourth year taking the place of the Sentence Completion Test, show that both of these tests were measurably different from those for which they were substituted.\*

*The Cancellation Test* was carried out in the simplest way—crossing a single letter from a mass of pied letters of the alphabet (using the Whipple

---

\* The Mutilated Text accuracy correlated with the Sentence Ideas of the first year to the extent of only .19, while the Mutilated Text speed index (number of ideas per second) correlated with sentence index to the extent of .26. The Cause and Effect test was more closely related to the Opposites test than the above two tests seem to have been. Accuracy of Cause and Effect with Accuracy of Opposites correlated .27, while speed of Opposites with Cause and Effect speed showed a correlation of .37. Because of the low correlations of consistency it seemed best not to include records of the Mutilated Text test or the Cause and Effect test in the later reckoning.

small letter form). Two measures were used in this test, the accuracy or percentage marked, and the speed index, derived by dividing the time by the accuracy. The irregularity in the succession of scores presented below was due to the fact that different letters were crossed on different years, "a" on the first year, "m" the second year, "w" the third, and "a" again on the fourth year. This was to avoid the possibility of practice affecting the scores of certain individuals.

The numbers below refer to the arithmetic mean and the standard deviation of each of the test measurements during the four years of the experiment on the 203 individuals. The standard deviations are in parenthesis in each case.

	1st Year	2nd Year	3rd Year	4th Year
Can. Acc.....	80.67 (15.21)	93.88 ( 5.56)	90.37 ( 7.69)	92.30 ( 7.91)
Can. Ind.....	24.33 ( 6.64)	18.45 ( 3.55)	19.54 ( 4.46)	19.93 ( 5.43)

The particular *Substitution Test* devised and used by Mrs. Woolley is difficult to describe adequately. For those who care to use it, the previously mentioned monograph should be referred to in detail. The main difference between the Woolley Substitution Test and other Substitution tests in frequent use is in the last retention page of the Woolley type. Not only is the time and accuracy recorded for the

learning pages (in each of which number-symbol substitutions are made with the key before one), but also on the final or retention page, in which case the key is removed and the subject recalls the individual substitutions from memory. It should be added that in no case is the subject able to make substitutions in a routine way by referring to previously recorded writing of his own, as each line of substitutions when completed is covered by the experimenter. It turned out that all the measures of the Substitution Test correlated positively and quite highly together, with the exception of the learning speed\* and the accuracy of retention.

Subst.—	1st Year	2nd Year	3rd Year	4th Year
(Learning Speed)...	1.426 (.301)	1.318 (.274)	1.277† (.271)	1.291† (.266)
(Retention Speed)...	1.307 (.636)	1.185 (.599)	1.284 (.680)	1.227 (.731)
(Retention Accuracy)	92.21 (12.10)	93.06 (12.83)	90.40 (13.10)	92.79 (11.60)

*The Memory Test* was one in which the subject was shown two seven-place numbers, two eight-place and two nine-place numbers, and was asked to read these different series out loud with the experimenter at the rate of one second per digit, recording them immediately afterwards. At first the

---

\* The learning speed is obtained by dividing total time in seconds by percentage correct.

† On the third and fourth years, two pages instead of three were given as a basis of learning.

scores for the different lengths of digits were kept separate, but later, on account of the individual fluctuation of separate scores, it was decided to pool all the results together, and deal with an average percentage of accuracy of all six series digits. We tried other measurements in connection with this test. Especially were we interested in finding out whether subjects who were variable as to accuracy in one year would also be variable in the next. In other words, have we here a reliable index for the "capacity for resisting distraction" which is measurable in terms of the amount of mean variation in the scores of the memory series? Our findings were negative. In 100 random cases a foot-rule  $R$  of .11 was all that was evident between the mean variation of one year and the next. The span of memory, the longest series of digits recalled, was also correlated in these 100 cases for the first two years, and a much lower reliability index was found than in the case of average percentage. So we finally considered only the one measurement of rote memory, and held to that—the average percentage of numbers recalled out of six cards read out loud by experimenter and subject together.

	1st Year	2nd Year	3rd Year	4th Year
Memory Accuracy.....	76.13	80.47	84.83	85.27
	(13.90)	(13.80)	(10.80)	(11.47)

Our *Sentence Completion Test* was the type suggested by Binet originally. A series of beginnings of sentences was presented, and the subject was

asked to complete each sentence. Various methods of scoring were arranged for by Mrs. Woolley. The types of measurements tried out in correlating were (1), the number of grammatically correct sentences written (referred to later as number O. K.); (2), the speed of association, measured by the number of sentences begun within two seconds' time after exposure of their beginnings; (3), the number of different ideas written down by the subject in the entire blank of thirteen sentences; (4), the speed index, or number of seconds per each idea written. The first two measurements were dropped from a good deal of the later manipulation of correlations, mainly because the usage of different sentence blanks of varying degrees of difficulty lowered their reliability correlations (between one year and the next) too greatly.

	1st Year	2nd Year	3rd Year	4th Year
Sentences				
(Number O. K.)....	11.11	12.10	11.78	
	( 1.94)	( 1.28)	( 1.67)	
Assoc. Speed.....	5.59	6.11	5.33	
	( 3.62)	( 3.82)	( 3.62)	
Number of Ideas.....	18.60	23.29	22.52	
	( 5.95)	( 6.78)	( 6.75)	
Speed Index.....	12.51	11.47	12.25	
	(5.516)	( 4.88)	( 5.91)	

*The Opposites Test* was used on all years except the third. The first two years we used rather easy blanks, whereas in the fourth year we used blanks of difficult opposites, made up largely of words



given in Simpson's lists of hard opposites. There was some indication of a difference in the mental capacity utilized for the easy lists as contrasted with the capacity necessary in attacking the more difficult lists of words. It was probably due, however, to the emotionally discouraging effect of the hard list on many subjects. Several persons who had done fair work with the easy list had to be repeatedly coaxed before even attempting to go over the list of hard words. In every case with the Opposites tests, two measurements were considered—the percentage of accuracy and the speed index, or the time divided by accuracy. For most purposes it was decided that the measure of accuracy better represented the test as a whole. The speed index showed a smaller reliability from one year to the next.

		1st Year	2nd Year	3rd Year	4th Year
Opposites	Accuracy....	77.45	78.72		51.43
		(16.30)	(16.91)		(22.74)
Opposites	Index.....	1.484	1.758		9.836
		(.664)	(.951)		(10.89)

There were two other types of measurements which we have used to a large extent, in addition to the measurements of the individual tests. The first was an average test standing of an individual in all the tests on a given year, referred to as "yearly average." This was found by averaging the numbers standing for the decile divisions into which each of the important test measurements of



an individual fell on a given year. (The system of describing a measurement by giving it a number, from one to ten inclusive, depending on the standing of the individual in comparison to the total number of subjects, is explained in full in the Woolley and Fischer Monograph.) Also there is a "total-test-average," the average of all these yearly totals. This final average, covering 57 different test measurements from four different yearly testings, is to our mind a fairly adequate statement of mental intelligence, so far as this can be measured by our standard tests. Trying to keep in mind the limits of the truth of this comparison, we have used our total-test-average quite extensively for the purpose of determining the possible change in the relationship of each specific test measurement to general intelligence from year to year.

A second type of measurement which we used quite extensively is that of the school grade which the subjects had completed at the time of leaving school. It has already been stated that the boys tested had completed at least the fifth grade of the Cincinnati schools. The percentage of those completing the fifth, sixth, seventh and eighth grades turned out to be 29, 30, 26, and 15 respectively. Although the distribution is not scattered enough for the purpose of getting very significant correlation indices between the amount of schooling com-

pleted and other functions, yet some interesting tendencies are suggested by such statistical treatment.

*The Method of Correlation* used almost exclusively in this research is that of getting the simple product-moment correlation index between the different functions concerned.

$r = \frac{\Sigma(xy)}{n \sigma_x \sigma_y}$  We are

not interested in the extension of new mathematical devices, or in the use of old ones, such as the correction formulas of Spearman, which have been questioned by Brown and others. It is doubtful whether the method of attenuating correlations, on the plea that the various yearly records were merely samples of exactly the same mental trait, is legitimate. In most cases the reliability indices from one year to the next are too small. But granting the legitimacy of the method, Simpson ('12), Webb ('15) and others have questioned the value of the expenditure of time involved in raising the average correlation index five or ten points. We believe that the raw correlation indices furnish us sufficiently accurate information upon the relationships with which we are interested. It is quite certain that the correlation ratio index,  $\eta$  would have given us higher

results throughout. As with Brown, all correlations between a speed and an accuracy measurement were slightly "j'd" in their plotting. Notwithstanding the higher set of indices that would have accrued by using the correlation ratio throughout, we feel that the extra time involved in this method was not

compensated for sufficiently. So we have concerned ourselves with linear regression measurements only—the product-moment “ $r$ ” index.

In the end we have gone somewhat beyond the bounds of the simple product-moment formula in attempting to present a number of partial correlations—correlations between mental tests stripped of the influence of school-grade-completed. In this connection we have utilized the formula for multiple correlation introduced by Yule and used in Psychology by Brown and Kelley.



## SECTION IV.

### RESULTS.

#### *Influence of age and experience on correlations between the same tests on different years*

After stating the purpose of the research, and describing the subjects and tests, it seems only necessary to add the tables of product-moment correlations as procured. Tables I and II have to do with the influence of age and experience on correlations between samples of the same tested capacity on different years. Are tests, e. g., of immediate memory, much more closely correlated when separated by short intervals of time, say one year, than when separated by longer intervals of two and three years? Between samples of the same test, we would hardly expect a closer correlation in the case of long-time intervals than in short-time intervals.

Table I refers to the correlation between the general test averages of the different years, and also the correlation of the different year averages with the total average. The method of getting these averages has been briefly described. It should be noted, however, that there is a difference between the method used in getting the fourth year averages and that used in the case of the other three yearly averages. In the fourth year, instead of assigning each test record of each individual to a decile division (designated from one to ten) on the basis of the records of all the subjects tested that year—between 350 and 400—as was done by Mrs. Woolley,

the individual records were assigned to separate decile divisions on the basis of only our 203 tested persons. This was done because records of all subjects were not easily accessible to the author in the latter part of this investigation.

*Table I* (showing the relationships between total test average and the averages of different years).

	1st Year	2nd Year	3rd Year	4th Year
Total Average.....	.89	.89	.85	.91
1st Year Average.....		.74	.69	.76
2nd Year.....	.74		.71	.76
3rd Year.....	.69	.71		.73
4th Year.....	.76	.76	.73	

The probable error of the correlations in the first row is approximately .01, and the probable error for correlations in the neighborhood of .70 is .024. In every case the number of individuals tested was 203.

Apparently there is little variation in the amount of correlation between the total test average and the different averages. It cannot be said whether the initial, or one of the later series of tests, conforms more closely to actual mental ability. The slight deviations in the third and fourth year correlations can be explained, largely, no doubt, by the fact that certain of the special tests helped to make up the averages, but were not included in our report because they had no duplicates in other years.

There is a marked consistency of fidelity to type of individuals over long-time periods, as shown by the intercorrelations of yearly averages.



Evidently the fourth year average was slightly closer to a reliable mental test statement than the other three year averages. This may be due to the different method used in computing this year's average, as already explained. This is shown by the high indices where (1) and (2) are matched with (4), as contrasted with the indices of (1) and (2) correlated with (3).

But the significant thing is that the correlations of (4) with (1) stand out as high as, or higher than, the correlations of (4) with (2) and with (3). As we shall discuss later in full, we feel bound to conclude that the factor of general test ability is so persistent among the individuals that age and experience do not interfere markedly with their relative position.

The next table has to do with the correlations between samples of approximately the same test measurements on different years. In Table II we are interested to know first, what reliability have different tests from one year to the next; i. e., how close is the correlation in the case of adjacent years? Secondly, what stability do test measurements have over longer time intervals? If the correlation between years which are not adjacent is markedly less than the correlation on adjacent years, we feel justified in concluding that this test ability tends to change over long-time periods. Whereas, if the correlations in the case of longer intervals tend to

be nearly as great as the correlations between adjacent years, we would conclude that the capacity does not change markedly on account of age and experience.

*Table II* (Correlations between samples of the same tested capacities on different years).

Samples separated by:

	1-Year Intervals				2-Year Intervals		A 3-Year Interval
	(1)(2)	(2)(3)	(3)(4)	(3)(4)	(1)(3)	(2)(4)	(1)(4)
Cancel. Accuracy .....	.22	.37	.33		.19	.20	.04
Cancel. Speed Index .....	.41	.54	.58		.43	.49	.50
Subst. Test Speed of Learn. ....	.60	.65	.81		.50	.54	.49
Retention Page, Speed .....	.46	.43	.59		.41	.46	.20
Retention Page, Accuracy ....	.50	.52	.47		.48	.47	.52
Immed. Mem. for Numbers.	.61	.62	.63		.60	.55	.62
Sentence Compl. Test — Number Correct Sentences ...	.47	.43			.34		
Assoc'n Speed..	.29	.42			.39		
Number, Diff. Ideas .....	.47	.49			.47		
Sentence Speed Index .....	.57	.50			.35		
Opposites Test Per Cent. Accur. ....	.52					.49	.43
Speed Index...	.44					.33	.43

With the exception of the Opposites test indices, the probable error varies between .043 for indices in the neighborhood of .30 to .030 for indices in the neighborhood of .60. The number tested was within five of 200 in every case. In the case of the Opposites Test indices, where 100 or fewer subjects were used, the probable errors are somewhat higher, varying from .051 for the indices around .30 to .050 in the case of the index of .52. For a complete table of probable errors usable in testing the validity of our indices, refer to the appendix.

The above results show features of great interest, both in regard to the specific tests in question, and also in regard to the general tendency of different types of tests to change from one year to the next. It may be of value to consider each test separately at first.

At a glance it can be seen that the speed index is more reliable than the accuracy as a test measurement of the Cancellation Test. It might be argued that this is due to the fact that some letters are seen with greater ease than others. On this account, accuracy would be a real factor of intellectual discrimination in one case, while it would not in others. Only carelessness would cause errors in the "m" cancellation test, for example. That the letters are probably not important factors in the result is shown by the fact that there are no high correlations anywhere, and also that the first and the fourth

years' results are so strikingly far apart ( $r = .04$ ), in spite of the fact that "a" was cancelled in both these years. In the case of the Cancellation speed index, however, the correspondence is remarkably close for the first and fourth years ( $r = .50$ ), where the letter cancelled was the same for both years. The speed index in the Cancellation Test seems to be a stable test measurement throughout, changing very little from one year to the next. That the speed index is so much more constant in this test than the factor of accuracy is, we believe, not due to the fact that people tend to be assigned more permanently in terms of speed than of accuracy. This is disproved by the results of other tests. We conclude, as suggested above, that the phenomenon is due to a difference in the attitude towards the test. At fourteen years of age, when the Cancellation Test is first attacked, the factor of intellectual foresight is really important and prominent, whereas in the later tests, particularly the test given in the fourth year, the function tends to become automatic, and errors are due to a carelessness of another type. But at neither time is there a close correspondence between accuracy of Cancellation and total intelligence.

The Substitution Test shows results entirely different, so far as the factors of speed and accuracy are concerned. With the learning pages, and also the last retention page, the speed of doing the test

gives much higher reliability coefficients when adjacent years are considered than is the case when longer periods intervene between tests. The accuracy of the retention page, on the contrary, shows practically no variation so far as the spread over varying lengths of time is concerned. The "r's" for longer intervals are as high as those for adjacent years. This may be due to the fact that accuracy of the sort required in such work does not tend to shift much among individuals during long periods of time. Or, more probably, it is due to the fact that the memory aspect of this test is the stabilizing influence which counteracts time discrepancies in accuracy of work, and consequently brings the long interval correlations up to such a high point. At least we must conclude, in this particular test, that although for the adjacent years or short interval periods, the correlations are much higher in the case of the speed of preliminary learning than with accuracy of retention, yet, when referring to the stability of individuals over longer intervals, accuracy correlations are just as high as those of speed. There is evidently a greater error in the individual measurements of accuracy in retention than in the measurement of speed. But the amount of change in individuals due to age and experience is not nearly so great in the former measurement.

The Memory Test shows a result very similar to that of the accuracy of the Substitution Retention



Page. There is no strikingly close correspondence between individuals from one year to the next. The correlations of adjacent years is never much over .60. But there is a marked faithfulness to type on the part of individuals over the whole period of four years, greater than in any of the other tests. In other words, the important factor in keeping down all correlations is apparently the sum of many chance disturbing factors, such as inattention, daily variation, auditory and ideational distractions, which enter into individual test performances to hinder the procuring of ideal scores of ability. The change in the individual from one year to another in this rote memory is not affected strongly by the factors of age and experience.

Because it was not used the fourth year, the Sentence Completion Test did not give as complete a set of correlations as did the preceding tests. In this test the first two measurements tried out are clearly not as reliable as the measurements of the number of ideas written and the speed index. As would be expected, age and experience evidently have an influence on the factor of the number of sentences written correctly, whereas this is not clearly true in the case of speed of association (determined by noting the number of sentences begun without pausing longer than two seconds). But in both of these measurements, the peculiarities of the particular test blanks which were used, and

other disturbing factors previously mentioned, seemed to play too important a part in reducing the correlations of reliability. For this reason we have not considered these two measurements in most of our later correlation tables.

Regarding the number of different ideas written in the Sentence Completion Test and the speed index, the results seem to point to the same type of conclusions to which we came in the case of the Substitution Test measurements. The measurement concerned with the thinking up of a large variety of ideas, apparently a sort of free association and certainly closely connected with a certain type of memory, is highly stable over long periods compared to the ability to think of these ideas in the shortest possible time.

The Opposites Test does not present results clearly in line with the other findings. In this test, in the case of long intervals, there is an apparent tendency towards a closer fidelity among speed measurements than among those of accuracy. We find it difficult to account for this in any way. The high correlation of the speed index between the first and fourth years might, however, be due to the fact that the first year and the fourth represented the more serious attempts, whereas the second year's blank was enough like the first to afford a let-down for many who did vigorous, alert work the first time. The fourth year's Opposites Test was, on

the other hand, difficult enough to call for the best in every one. Again the high stability of the speed factor, compared with the accuracy measurement, might be explained on the basis of the change in the type of test. As was said before, the fourth year Opposites Test was a quite different affair from the Opposites Test of the first two years. The distribution of the subjects was around a far lower percentage value; and there were a number of persons who had been quite accurate during the first year, but who fell down badly on account of emotional causes when confronted with the harder list.

SUMMARY. (1) There are two important factors to consider regarding the reliability of a test-measurement over an interval of time. In the first place, we want to know whether the test is a reliable one, bringing high correlations between its samples, over short intervals of time. In other words, is the measurement one which can be closely duplicated shortly afterwards? The second factor to consider is whether, regardless of the amount of reliability of the measurement itself, there is a high stability in the tested capacity over long-time intervals.

(2) On the basis of the above results, it appears that those tests which have memory as an important item in their make-up, whether immediate or secondary memory, are of the type difficult to measure accurately at any one time on account of the disturb-

ing factors which enter into the test procedure. But these same tests measure capacities which are very stable over long periods of time. Those who have good memories, immediate or secondary, appear to hold faithfully over a period of three years to their relative positions in the group.

(3) Those test measurements included in our series of tests, which have to do with speed in work, appear to be less influenced by the disturbing factors in the test administration than are the memory measurements, and frequently show a high reliability correlation over short-time periods. But individuals do not hold as faithfully to type in the case of speed measurements over long periods of time.

(4) In the case of accuracy in mental work, the results are not nearly so clear. Apparently with routine accuracy (as in the Cancellation Test) there is practically no faithfulness to type over long-time intervals. With accuracy of a high type, involving memory and associational factors, this conclusion seems not to be so valid. Our tests are not extensive enough to warrant definite statements in this report.

*The influence of Age and Experience on the Relationship between the Total-Test-Average and the Different Test Measurements.*

We have already referred extensively to the factor of Total-test-average, which is the average of the test performances covering a period of four



years. This is the closest approach we have to a statement of general intelligence, although it is obvious that such a total average will be partial in its inclusion of certain mental capacities and its neglect of others. For example, at least three speed records of the substitution test were included each year, but no measurements referring to accuracy of retention in this test. Yet, granting the onesidedness of this total estimate of intelligence, it will probably be serviceable to indicate in every case any marked change in the way a test is associated with intelligence from one year to the next.

Our question, then, is: Do the tests individually or as a whole tend to be correlated more closely to the total-test-average at the time of their first trial or later? Does previous familiarity with a test, and intervening age and experience, tend to make that test more, or less, closely related to general-test-intelligence? Table III presents this comparison.

*Table III.* (Correlations between Total Test Average and Individual Test Measurement.)

	1st Year	2nd Year	3rd Year	4th Year
Cancellation Accuracy	.34			.38
Cancellation Speed Ind.	.49	.31	.43	.42
Subst. Test Retent'n				
Accuracy .....	.22	.22	.28	.30
Subst. Test Speed of				
Learning .....	.65	.65	.56	.54
Immediate Mem. Test	.70	.63	.56	.68
Sent. Compl. Number				
of Ideas .....	.36	.34	.24	
Sent. Compl. Speed Ind.	.49	.55	.53	
Opposites Accuracy...	.18	.35		.51
Opposites Speed Ind..	.45	.42		.45



The number tested was approximately 200 in every case, with the exception of the Opposites Test on the first year, in which case N was slightly less than 100. (Refer to appendix for a statement of probable errors.)

There is no general tendency for all tests to become either more or less closely associated with total-test-intelligence. It depends entirely on the individual test, its appropriateness as a real mental measurement in different years.

*The Cancellation Test* shows an interesting result, especially when compared with the results presented in Table II. It appears that the measurement of accuracy is slightly more closely related with general intelligence at the end of three years than initially, despite the poor stability of the measurement from year to year. The speed index, however, in which was found a high faithfulness to type among the subjects, shows a tendency for a drop in the degree of its relationship with total-test-intelligence in each succeeding year. In fact, in the fourth series of tests, it is almost as valid to use the percentage of accuracy as the speed index in picking out the more desirable subjects.

The correlations between the Substitution Learning pages and the total-test-intelligence decrease from year to year. This suggests the possibility for all speed indices to become less and less important with repetitions of the tests. The amount

of the drop in this case is so striking that it reveals again the general fickleness of this test in measuring capacity over long interval periods. Evidently those who are particularly good at the time of the first testing are frequently surpassed at a later time by those poorer in general test ability. In the case of the retention accuracy, there is an evident rise in its relationship to the total-test-average. This may be due entirely to the tendency of some subjects, while going over the learning pages in the later years, to anticipate more definitely the final retention page. But on account of the consistency of the increase in this correlation, even from the third to the fourth year series, we might well be justified in concluding that care, such as is called for in our Substitution Retention Page, does become more important as an index of intelligence from year to year.

In the Memory Test, there is a sameness in the correlations from one year to the next, with the possible exception of the correlation on the third year. This is especially significant considering that the scores in this particular rote memory test (see page 34) approach more and more to 100 per cent. accuracy. If we had a harder memory test, which would differentiate more exactly the good from the mediocre, we would probably find that our memory measurement had become even more closely related to total-test-intelligence on account of age

and experience. As it was, on the fourth year of the test as many as 20 got perfect scores, while on the first year only four attained 100 per cent.

The results with the Sentence Completion Test, apparently upset any assertion we might feel justified in making as to the general yearly decrease in the relationship between speed test and total-test-intelligence, and a corresponding increase in the relationship of accuracy tests to test intelligence. The "number of ideas written" might well come under the heading of mental accuracy, and yet there is a marked fall from the first to the third year in the relationship between this measurement and the total test accuracy. The nature of the test, however, will account for a large amount of this drop. The test allows for so much freedom of association, since subjects are not told explicitly to write as involved sentences as possible, that it was evident at the time the tests were given that many of the brighter boys did not write as many ideas in the later years as initially. On the other hand, they had been timed directly and urged to hurry up so much in the other tests that, although the timing of the Sentence-Completion Test was intended to be without the subject's knowledge, they did get the notion of speed rather than of fullness in what they wrote. The fact that the better subjects did speed up in their writing, on account of the influence of the other time-taking tests, has operated to make

this speed index more important from year to year. In fact the third year speed index results are about as close as any of the tests to total-test-average.

The Opposites Test shows a rise in the relationship between accuracy and total-test-average that is very striking. This is clearly indicative of the quite different nature of the fourth year's difficult blanks as contrasted with those blanks used on the first year. The words used on the fourth year necessitated a certain amount of understanding to appreciate their bare meaning, while the first year's easy Opposites blanks presented no difficulties to the subject so far as grasping the meaning of the words was concerned. The hard Opposites list measures not only the capacity to write the correct Opposites, but the ability to face with self-assurance a very difficult task. On the whole, then, the only conclusion we can make regarding the Opposites test is to the effect that a really difficult blank of test-words challenges mental test ability much better than the easy blanks.

SUMMARY. We can say, then, that there is not a marked tendency, in the case of most of the measures, towards either a greater or a smaller correlation with total-test-intelligence on account of age and experience. It is quite certain that no general statement can be made that will hold good for tests of all kinds. As concluded by Brown, Abelson, Desourdes and others, those tests which



measure mental ability adequately among subjects of one grade of intelligence or age do not fit in well as measures of ability in subjects of a different sort. At the age of fourteen, and when unsophisticated by previous mental tests, speed in work is very essential, and accuracy less so; whereas, in the latter years, speed is a less important factor, and the ability to be accurate, and especially to face unusual and difficult tasks composedly, is of greater significance. Hart and Spearman ('14), on the basis of a large number of tests on insane adults, and also from their contact with results of tests on normals, conclude that accuracy is a more important mental trait than speed. We would supplement this by the statement that at least accuracy seems to become more important than speed, in measuring general intelligence as people grow older.

*The Change in the Correlations Between Different Tests From One Year to the Next.*

We have presented tables of correlations estimated to show the change in the validity and meaning of measures from certain standard mental tests, which change was affected by the factors of age and experience. This was done in part by showing the differences in the reliability of these measurements when they were tested out with short and with long intervening time periods, and in part by comparing the relationship between certain of these measurements and total-test-intelligence in the first year of



their administration as well as later. Our next endeavor will be to find out the change in correlations between different tests from one year to the next. As stated in the introduction, we are especially interested in finding out if tests are more or less closely correlated with each other at the time of their first trial, as compared with later administrations of the same series. After a survey of the results of the preceding sections, it appears doubtful whether we can get at any conclusion on this question. Many of the tests have apparently changed in their function as mental tests enough to invalidate strict comparisons from one year to the next. At least it is important that we take into account these changes indicated in Tables II and III, in the case of each test's correlations.

Only those tests with a fairly high degree of stability were used throughout in these comparisons. Otherwise we would be entirely unable to interpret results. To do this we have limited ourselves to those tests which gave reliability correlations of .40 or higher. Three exceptions, however, were made to this minimum reliability limit, the Substitution retention page speed, and the "number of correct sentences" and "speed index" of the Sentence Completion Test. From our five types of tests, repeated on at least three of the years, eight measurements were chosen for a comparison of the inter-correlations from one year to the next. On

account of the substitutions already referred to, there was an omission of all correlations connected with the Opposites Test on the third year, and the Sentence Completion Test on the fourth year. A complete report of the correlations between tests of different kinds is found in Table IV.

We must admit that our data are quite inadequate to answer our first main question. We cannot say whether or not tests as a whole become more closely related from one year to the next. So many irregularities seem to be present throughout the course of the four years, that we are at a loss in even attempting to generalize in the case of many individual pairs of test correlations.

In our set of ten groups of four indices each (concerning those measurements used throughout the four years), there is an unequivocal rise in only two of the ten groups—Substitution learning speed with Immediate Memory, and the accuracy with speed of the retention page of the Substitution Test. The later of these two rises may be said to be due to a lessening of the factor of accuracy from one year to the next. The former is difficult to explain without some general intelligence hypothesis (as suggested by Hollingworth), and the assumption of an increase in correlation due to age and experience, including practice. But this explanation would have more weight if borne up by results from other test measurements.

Table IV. *Correlations between different tests on each of the four years:*

		Substitution				Sentence Completion						
		Canc. Index	Learn'g Speed	Ret. Speed	Ret. Accuracy	Immediate Mem.	No. O. K.		Opposites Acc.			
							Speed	Ideas				
Canc. Sp. Index	{	.27	.23	.16	.25	.16	.07	.12	-.10			
		.33	.04	-.02	.22	.25	.02	.15	-.18			
		.39	.13	.07	.25	.26	.08	.35				
		.38	.22	.20	.27				-.06			
Substitution	{	{	Learning	.27		.47	.31	.25	.10	.07	.10	.11
			Speed....	.33		.53	.37	.07	.07	.15	.23	.19
				.39		.54	.37	.28	.04	.10	.43	.24
				.38		.45	.31	.43				.24
	{	{	Retention	.23	.47		.71	.22	.03	.01	.09	.02
			Speed....	.04	.53		.73	.13	-.01	.04	.03	.13
				.13	.54		.77	.14	.09	.02	.14	
				.22	.45		.82	.19				.03
	{	{	Retention	.16	.31	.71		.14	.12	.04	-.01	.09
			Accuracy..	-.02	.37	.73		.10	.12	.05	-.10	.16
				.07	.37	.77		.17	.13	.10	.03	
				.20	.31	.82		.11				.06
	{	{	Immediate Mem.	.25	.25	.22	.14		.27	.18	.29	.25
				.22	.07	.13	.10		.24	.05	.26	.18
				.25	.28	.14	.17		.08	.07	.18	
				.27	.43	.19	.11					.31
Sentence Completion	{	{	No. O. K...	.16	.10	.03	.12	.27		.61	.45	.09
				.25	.07	-.01	.12	.24		.50	.48	.23
				.26	.04	.09	.13	.08		.57	.44	
												.24
	{	{	Ideas .....	.07	.07	.01	.04	.18	.61		.38	.24
				.02	.15	.04	.05	.05	.50		.36	.13
				.08	.10	.02	.10	.07	.57		.35	
												.13
	{	{	Speed Index	.12	.10	.09	-.01	.29	.45	.38		.13
				.15	.23	.03	-.10	.26	.48	.36		.18
				.35	.43	.14	-.03	.18	.44	.35		
												.13
	{	{	Opposites Acc..	-.10	.11	.02	.09	.25	.09	.24	.13	
				-.18	.19	.13	.16	.18	.23	.13	.18	
				-.06	.24	.03	.06	.31	.24	.13		

For reference to the probable errors of these indices see Appendix. Those correlations concerned with the first and second years' Opposites test were made on the basis of 100 individuals, or slightly less. All other correlations refer to cases in which "n" was approximately 200.

The remainder of these ten sets of correlations give great irregularities from one year to the next, or else show no tendency either to increase or decrease on account of age and experience. Evidently the variations in the Cancellation Test, different letters being crossed out on different years, have interfered with consistency in the correlations connected with this measurement. It is hard to see why Substitution retention accuracy does not become more and more like the Immediate Memory measurement. This seems to support Wyatt's ('14) conclusions as to the distinctness of certain memory measurements. Our results are especially important since continued experience does not lessen the amount of disparity between the memory tests. There appear to be few cases of consistent drop in the amount of correlation between test-measurements continued throughout the four years or with those tests administered on three of the four years. The striking exception to this is the drop in the relationship between Immediate Memory and the Completion Test speed index from one year to the next. It seems odd that this should be the case, especially as we have already noted the growing close relationship of Completion Test speed with



total-test-intelligence from the first to the third year. It is one of the marked discrepancies which would have to be faced by those who favor the theory of the general intelligence factor and an increase in correlation between tests in the case of practice. It looks to us as though this particular set of correlations, and also some of the less regular drops in the amount of correlation between tests from one year to the next, strongly intimate an actual levelling process in mental ability in the case of the Cincinnati subjects. It is certainly not improbable that as people advance in years from fourteen to eighteen there is a compensating vocational influence present in such a way that those subjects who have specialized in the use of one type of mental ability are apt to fail in another type.

In a number of the groups of correlations, we note one particular inclination of great interest—the fact that many correlations seem to drop from the first to the second year, and rise again slowly from the second year on. This suggests to us an hypothesis which we believe clarifies somewhat the marked divergence between the theory of Binet and Burt on the one hand, and on the other the belief of Spearman and Hollingworth that practice always tends to increase the breach of difference between people, as shown by rising correlations between practiced test measurements. In accordance with the suggestion of Binet and of more recent experi-



mentalists, we believe that original ability to adapt to new situations, and to understand the instructions such as are given, is enough of a single trait in itself to raise the correlation between two performances beyond what would be the case if the instructions were well understood from the beginning by everyone, and where each subject was ready to do his best. Obviously, as the understanding of instructions does enter into all tests, this common permeating influence would help to separate individuals into good or bad in all tests alike. This would operate to raise the amount of correlation on the first year to a higher degree than if the understanding of the instructions did not enter into the situation. In the second year, the instructions are presumably well understood, as the tests are no longer new to the subject. Then, just because a person is quick in understanding instructions does not mean that he will be capable in all the tests.

We would include as the second part of our hypothesis the expectation of a steady rise in test correlations after the first year. This is partly because, as suggested by Hollingworth, the tests become more alike due to practice, and partly because of the greater predominance of the common factor of general intelligence under the conditions of practice. If we take the average correlation in the case of those tests which were correlated on the first three years together, we find that the aver-

age correlation made in connection with each of the tests of importance (omitting the Opposites test), stands as follows:

*Table V.* (The average correlations of each test measurement with every other test measurement on each of the first three years of testing, Opposites Test excepted.)

	1st Year	2nd Year	3rd Year
Cancell'n Index.....	.18	.14	.21
Subst. Learning Speed.....	.22	.25	.29
Subst. Retention Speed.....	.25	.21	.28
Subst. Retention Accuracy.....	.21	.18	.22
Immed. Memory.....	.23	.15	.21
Sentence Test: No. O. K.....	.25	.24	.26
Sentence Test: No. Ideas.....	.19	.17	.20
Sentence Test: Speed Index...	.21	.20	.27
Average intercorrelations of all the tests.....	.217	.193	.242

By inspection it is clear that the mean variations of each of these averages is so great that direct comparisons between years in the case of individual tests is not very valuable. The uniform change in amount of correlation for many of the tests is however significant.

The individual test measurements present some, variations of importance. The Sentence Test measurements, for example, conform less to the average rule of rise in correlations after the first year than do the other measurements. This is probably due largely to the change in the quality of the test itself from one year to the next, as noted on page 55. The rise in the correlations, in the case

of those tests which are of the speed variety, is not great except when the speed measurements of the first three kinds of tests are correlated with the speed index of the Sentence Completion Test. Again the change in the type of the test is responsible. The practical absence of an alteration in the amount of correlations between our two types of accuracy-memory tests has already been commented upon. It seems, then, as though the rise in correlations was not due to a greater sameness in those tests which appeared to be most alike, but rather to an increase in the amount of correlation of those tests which are apparently quite different from each other. Whatever specialization among Cincinnati boys has taken place from one year to the next has apparently been a specialization in *content*, rather than in a psychological *process*. That is, a good memory in one type of material does not mean an efficient memory of another sort, and speed in one test does not correspond to speed in another. So far as other types of less closely related measurements are concerned, such as speed with an accuracy measurement, there appears to be steady increase from the second to the third and fourth years. Whether this is due to the factor of practice entirely, or to the tendency for boys to separate themselves more distinctly under the Cincinnati conditions on account of age and experience, we are not ready to say definitely. We are inclined to believe the latter.

SUMMARY. (1) It appears that there is no marked universal tendency for correlations of test-measurements either to increase or to decrease on account of age and experience, under the conditions of the Cincinnati experiment. There are many striking irregularities which are difficult to explain in the course of the correlations from one year to the next.

(2) On the whole, it is evident that many tests decrease their correlations from the first to the second year, due presumably to the factor of the understanding of instructions (common to all tests at the time of their first administration). After this initial drop in the amount of correlation between the tests, there is likely to be a slight increase, on the following years, in the amount of correlation. Whether this is due to practice alone, or whether the factors of age and vocational experience varying widely among the subjects contribute to this, it is hard to say. We are inclined to believe that the vocational life in Cincinnati during these three years does aid to some extent in differentiating the good from the bad. This is in spite of the fact that, in the case of certain types of tests, there is an evident evening up process (as in the memory tests), so that those who are proficient in one are not correspondingly proficient in the other, even after practice. The increase in the amount of correlation between different tests is greater when the tests are apparently unlike each other than when the tests are alike.



*The Influence of Age and Experience on the Relationship Between Different Mental Tests and the Amount of School Attendance.*

This topic introduces us to a new measurement of ability and the whole question of the influence of school training on intellectual capacity. As before indicated, our group was made up of subjects taken from the fifth, sixth, seventh and eighth grades—that is, those who had passed in these grades.

In this connection the following questions arise: Which tests seem to be most closely related to the amount of school work procured? In the course of the growth of the subjects during the three years away from school, does the relationship between school grade and test-intelligence alter markedly? Which tests are related the most closely to the amount of schooling procured as indicated by the drop in their correlation with this function on account of age and experience?

As previously mentioned, other researches in the field of mental tests (especially Bonser's) have attempted to relate the results of mental tests with ability expressed in the different school subjects, but so far as we know no wholesale attempts have been made to correlate tests, or the average of a number of tests with the total amount of school work undertaken by subjects of the same age. The following indices are



the correlations between the school grade completed and the averages of the mental tests on the different years.

	1st Year	2nd Year	3rd Year	4th Year	Total Ave.
School Grade..	.51	.45	.51	.66	.63

The rise in the amount of correlation from the third to the fourth year has two sources of explanation, already noted above: (1) the difference in the tests used—Mutilated Text instead of Sentence Completion Text—in the fourth year, and (2), a change in the method of computing the mental average on the fourth year. The irregularity on the second year may be due in part to the influence of the puzzle box (Healy and Fernald), which was included in the summarizing of percentile averages in this year, a test which correlated well with nothing. Also, as suggested above, the understanding of instructions was an important item in the first year of testing, but less significant in the second year. In general it seems safe to conclude that the relationship of the various tests to school grade is not decreased on account of age and experience. If anything, there is a slight tendency towards a better correlation between the amount of school training and the results in the mental tests, even after three years of the influence of age and experience.

The correlation reached on the fourth year between school-grade-completed and the mental average as computed on that year seems, to us, remarkable. A good many explanations can be forwarded for this possible gravitation of those who were low in school ability to a relatively lower level of mental ability after leaving school, and a similar higher rating of those from the upper school grades. A few individuals who had finished the eighth grade, for instance, did get a chance to go into the night High School, while those who were not at the time through the grades, or who could not prepare for the High School work by taking one or two years of continuation school work, were usually not encouraged to do night school work. So far as further school work was concerned, there was an air of hopelessness about the boy who had completed only the fifth or sixth grade of work. But we think this factor was not very influential in the long run. Certainly not over five per cent. of the present 203 subjects took advantage of enough night High School studying to make a real difference.

The stimulus received from the higher school grades for more advanced reading and thinking, and the better grade of position taken by these seventh and eighth grade boys, were probably more significant factors in correlating the school-

grade-completed with test intelligence. Possibly the factor, reviewed and experimentally demonstrated by Thorndike ('16), of the greater improvability of the more intelligent persons, is of some influence. Those coming from the higher grades, and consequently of a better intellectual calibre, improve more through contact with the outside world than do those from the lower grades. The table of correlations between the school-grade-completed and certain of the mental tests follow below:

Table V. (Correlations of school-grade-completed with certain of the tests given on the first and later years).

	1st Year	2nd Year	3rd Year	4th Year
Cancellation Accuracy	.23			.21
Cancellation Speed...	.20	.05	.11	.26
Substitution Speed...	.21	.21	.25	.28
Substitution Retention				
Accuracy .....	.02	.07	.04	.17
Memory (Immed.)...	.47	.49	.49	.52
Sentence: No. Ideas..	.15	.20	.42	
Sentence: Speed Index	.44	.37	.33	
Opposites: Accuracy..	.10	.02		.11
Opposites: Speed Ind.	.33	.12		.30

There is little to add to the comments already made, except the more definite statement that in practically all tests there is as close a relationship with school-grade-completed in the third year after leaving school as immediately after. In many tests there is a drop in correlation on

the second year, due, no doubt, to the influence of adaptation to a new situation in the first year.

Three facts stand out as particularly important to us from a study of the table. In the first place, there is a high correlation between the factor of memory and the school-grade-completed. Also, these factors continue to correlate just as highly with age and experience. Apparently the training received from school experience in general, together with the better types of jobs taken by those boys who came out of the higher grades of school work, has acted in each successive year to maintain the relationship between the school-grade-completed and the capacity for rote memory as tested.

A second striking fact, and to us just as significant, is the low correlation between the factor of retentive ability and school-grade-completed. Either because the school has not trained children in this particular line of efficiency, or because intelligence in general does not rely much upon this trait in character, the correlation between school-grade-completed and retentive accuracy (as measured in Mrs. Woolley's Substitution Test) is practically nil. Only on the last year, on account of the influence of age and experience, does the relationship between school grade and retentive accuracy become at all significant.

The third item of considerable significance to us is the low relationship all the way through between the accuracy of the Opposites Test and the function of school-grade-completed. We would be led to expect a fair degree of relationship between these measures, at least as much as in the case of memory. The results are clearly to the contrary. Neither the ability to write down the Opposites to easy words at the age of fourteen, nor the ability to write down Opposites to hard words at eighteen, is related at all closely to the amount of schooling undertaken by children. This corresponds with Bonser's finding, that the Opposites Test, although superior to all other tests so far as measuring test intelligence is concerned, was below two of the tests so far as correlation with school grade is concerned.

As a general maxim, one might be led to conclude somewhat sweepingly that the school was giving too much importance to the factor of rapid and immediate memory work of a rote character, whereas the items of retention and of flexibility of ideational control were scarcely credited at all. We feel, however, that the problem is too complicated to make statements of such a dogmatic character, and that various types of investigations will be necessary for conviction on this point. Whether education, any other than that of a fairly specific kind, has any direct effect on



such a function as the ability to write opposites, has as yet not been proven. The degree of difficulty of the specific Opposites Test seems to have nothing to do with this situation, as the Hard Opposites Test correlated much more closely with total-test-intelligence than did the Easy Opposites Test, but no better with the school grade factor. We have also correlated the accuracy of the "cause and effect" Paired Associates Test with School-grade-completed, and find a considerably higher result than in the case of the Opposites Tests ( $r = .38$ ). But this test was clearly much more of a memory test, as given by us, than a test in controlled association.

SUMMARY: (1) The amount of school work undertaken is fairly well related to intellectual ability as determined by our mental tests. After three years of industrial experience, the relationship is as close as, if not closer than, it is when the children come straight from school.

(2) The relationship between school grade and individual test measurements remains remarkably stable for three years after leaving school. The amount of correlation in each case differs widely according to the type of measurement. There is a surprisingly high correlation between immediate memory and school grade, and a corresponding low correlation between our educational equipment factor and such measure-

ments as our retentive accuracy in the Substitution Test, and Opposites accuracy. Substitution learning speed, with its highest reliability correlations from one year to the next, has only a low correlation with school grade completed ( $r = .25$  on the average).

*The Influence of the Amount of Education on Correlations With Various Mental Tests.*

We are also interested in noting to what extent the amount of school work undertaken is an important factor in bringing about positive correlation between different tests, or samples of the same test. The mathematical device used in such a determination was first emphasized by Yule in his work on "The Theory of Statistics," and has been used in Psychology by Brown, Wyatt, Kelly and others. Following is Yule's notation:

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

where  $r_{12.3}$  stands for the correlation between the functions 1 and 2 with the function 3 constant, or ruled out.

The following table gives the straight correlations and the partial correlations (with school grade constant) for the first and final records in a number of tests:

	Years	School Grade Straight Constant
Cancellation Accuracy.....	(1) (4)	.04      -.01
Cancellation Index.....	(1) (4)	.50      .48
Substitution Speed.....	(1) (4)	.49      .46
Substitution Retention Ac- curacy .....	(1) (4)	.52      .53
Memory: Immediate Rote..	(1) (4)	.62      .50
Sentence: No of Ideas....	(1) (3)	.60      .61
Sentence: Index.....	(1) (3)	.35      .33
Opposites: Accuracy.....	(1) (4)	.43      .42

It is evident that the only straight correlations between tests influenced markedly by the factor of school grade are the correlations between samples of the Immediate Memory Test. In all other cases the relationship between the amount of schooling and the individual tests is so low that there is no marked change in the degree of relationship between tests when the factor of school-grade-completed is eliminated.

We are not interested in presenting complete tables of the intercorrelations of tests given on the same year with school grade constant, although we have computed many of these results. Only those correlations connected with immediate memory appear to be changed to any significant extent. In the first year's test records the correlation of the Substitution Test speed with Memory was reduced

from .25 to .18, when school grade was constant, and of Memory with Cancellation speed from .25 to .16. The correlations of Memory with retention accuracy, sentence index, and Opposites accuracy are, however, not influenced by keeping school grade constant. All other test measurements correlate so poorly with school grade that their inter-correlations are not materially altered by keeping the school factor constant.

*Summary Regarding the Characteristics of the Individual Tests Measured.*

We have taken up a fairly complete discussion of the results of each table of correlations at the time of their presentation. Possibly the best method of bringing certain results together in final form will be to discuss the characteristics of each individual test measurement.

The results from the Cancellation Test agree with inferences from Binet's work on practice in the case of cancellation, to the effect that accuracy is not as reliable a measure as speed.

But the apparent unreliability of the accuracy aspect of the test does not necessarily mean that, in picking out good subjects, the test is less valuable after experience than initially. The value of the test does not alter in its ability to signify intelligence. If anything, the speed factor tends to become less associated with total intelligence than ac-

curacy, with practice and experience, although it is a highly stable measurement over intervals of time.

In general, the results corroborate most of the other data on the Cancellation Test (reviewed by Whipple in his *Manual of Mental and Physical Tests*). Cancellation of a single letter has not, in any extended trial, proved to be highly valuable as a diagnostic expedient—particularly the accuracy measurement of such Cancellation.

The Substitution Test has no extended history, and the type used by Mrs. Woolley is quite different from any of the other varieties in common use. For that reason it is hard to make comparisons with other works. In cases, however, where a similar learning test has been tried there has not been a high correlation with imputed intelligence or with the records of other tests.\* It has not compared for instance with the Opposites Test, as a valuable mental measurement to use. Mrs. Woolley found it the poorest mental test, Cancellation Test excepted, in differentiating the grade groups in the fourteen and fifteen year old boys. Its value as a diagnostic test for vocational analysis is challenged by the author ('17), who tried it out on boys learning telegraphy. A much lower correlation with estimated telegraphing ability was found in the case of the Substitution Test than with the Memory or the Opposites tests. The results given above indicate a poor correspondence of the test with school-

---

\* See Whipple's *Manual*, p. 499, ff.



grade-completed, and a considerable degree of instability, as to the capacity measured, over long intervals of time. The speed index of the test, however, is a highly reliable measure for short-time intervals.

The Substitution Test retention accuracy is more like the Memory test than the Substitution speed index, to the extent that it is a relatively unreliable method for giving adequate single measures, but seems to test a capacity which is stable over long-time periods.

Our method of testing rote memory has not been duplicated by any experimenters who have done extensive correlating with mental tests. Sleight ('11), Wyatt ('14) and Carey ('15) have attempted to isolate a common memory factor from the factor of general intelligence by the method of partial correlations. The first two were unsuccessful, but Carey believes there was evidence for a slight special memory factor. Winch showed that, although the correlation between two memory tests may be low, there may be an association between the tests as evidenced by a transfer of practice from one test to another. We would hesitate less to say whether an immediate memory rote test, such as we have used, is able to represent some common memory function, if it were not for the fact that an entirely different type of memory measurement, represented by our Substitution retention accuracy, showed results similar to it.

Most experimenters who have tried out the reliability of their test measurements (Brown, Burt, Abelson, Simpson and others) have found only a mediocre reliability index in the case of memory measurements. In the case of stability of the memory, we find only one comment, a statement by Whitely to the effect that, in a test such as a memory test where the function has been frequently exercised as compared to capacities less frequently exercised, there is little change made on account of practice. This seems to fit in well with our findings. In general, we feel that there is a definable memory factor, and that it has the two characteristics, (1) that it is difficult to measure reliably on account of the chance conditions of the moment, and (2) that it is a markedly stable sort of factor in the life of each individual over a long period of time. Despite the difficulty of measuring the function reliably at any one time, it seems to correlate very highly with the amount of school work completed by the individual. It is interesting to record in this connection the comment made by Hart and Spearman, similar to a note of Burt's, to the effect that, in their opinion, the teacher's estimates of general intelligence were too highly colored by the ability of the child in rote memory.

The Sentence Completion Test was given only on the first three years of the testing, so that our conclusions are not as complete as in the case of

the other tests already referred to. On the whole, the speed of writing per idea seems to be the closest measure to our total test intelligence, although the mental function represented by it appears to be less stable over a period of years than the function represented by the number of ideas written.

We are convinced that the measurements connected with this test do change in their meaning when the test is repeated, possibly because there is no intimation of speed given in the instructions. At least it is true that the relationship between the different yearly samples of the same measurements and such stationary factors as school-grade-completed or total test intelligence, is considerably modified. This situation makes us undecided as to the relative stability of the various mental capacities measured. The fact that the "number of ideas" correlates with school standing more closely two years after leaving school than immediately after, suggests an interesting possibility. The inclination to continue writing long sentences when not told to do so is a characteristic of those with more school training, while the initial writing of long sentences is not so, to the same extent. The number of ideas written, however, becomes less important in measuring general intelligence with each repetition.

Objections have already been raised to the use of the two measures designated as "number of sentences written correctly" and "number of sentences written with pauses of less than two seconds." If

our sentence test blanks were more completely standardized, so that we could compensate for the difficulty of the harder blanks, the results might be different. However, the speed index, or number of ideas per minute, seems to be a highly valuable intelligence determiner, even without standardization.

The Opposites Test has been widely used by experimenters, and to good advantage. Bonser and Simpson both refer to it as their best single measurement of intellectual ability. And those authors using it, who have arranged their tests in order of a common factor of intelligence, have assigned it to a high position on their lists of tests. Our results corroborate Bonser's findings in that neither the easy nor the hard lists seemed to be closely related with school-grade-completed; whereas the hard list at least correlates highly with general intelligence. The factor of accuracy seems to be a more reliable single measurement than the speed index. It is no doubt a matter of attitude towards the test. One might be led to emphasize either speed or accuracy in the test, and the accuracy factor does not suffer as much as speed when relatively disregarded. The high correlation of the speed index on the first and fourth years is an anomaly hard to explain. Evidently the attitude of the first attack upon an easy list of opposites is more like the attitude towards a hard opposites list given for the first time three years later, than like the attitude towards a second easy list a year later.



## CONCLUSIONS.

I. There is a marked fidelity to intellectual type in individuals throughout the adolescent period of growth. A disagreement between the relative standings of subjects tested, year after year, is due to the chance factors of individual disposition and other incidents of our present testing methods, rather than to any striking change in the relative standing of subjects on account of age and varying experiences in the world of affairs. This is shown by comparing correlations between mental averages taken with long-time periods intervening as contrasted with similar correlations administered with short-time intervals.

II. The amount of school work completed also correlates well with average mental ability and, if anything, the correlation increases on account of age and experience in industry.

III. The reliability index of a test measurement (a correlation of samples of the same mental capacity repeated with short-time intervals) should be clearly differentiated from the stability index (a correlation of a similar type with long-time intervals). In general, speed of learning such as the measurement used in the Cincinnati Substitution Test, is a highly reliable measurement for the testing of its specific capacity at any particular time. But this capacity seems to be relatively unstable



over two or three year intervals. Workers who are rapid at one time may not be as rapid several years later. Our measurements of immediate and retentive memory are somewhat less reliable for any single testing, but they represent capacities which are highly stable over long intervals of time. Routine accuracy of the kind involved in the Cancellation Test is a measurement from which nothing can be inferred, two or three years after it is recorded. Measurements connected with such tests as the Sentence Completion Test and the Opposites Test are apt to change their meanings quite markedly over a period of years. This is due either to a change in attitude towards the tests or to a variation in the difficulty of the blanks used. There is no clear evidence that individuals would not be relatively true to type in all of these measurements over long periods.

IV. Immediate memory is correlated with the amount of school work completed, more highly than any of the other tests, and the amount of correlation does not fall off on account of age and experience. As compared with immediate memory, the speed tests are not nearly as closely related to the grade completed, and they become less associated on account of age and experience after the subject leaves school. There is a surprising lack of correlation between ability represented by the Opposites Test and the amount of school work undertaken. The difficulty of the blanks seems to make no difference in this respect.

V. On the basis of results procured for partial correlation with the usual formula, the amount of school work undertaken was not influential in bringing about any inter-correlations of test measurements. Some of the measurements connected with immediate memory are exceptions.

VI. We feel hesitant in drawing a conclusion regarding the tendency for people of adolescent age to become more or less alike as they grow older. A review of previous researches, and also a notation of the way our tests have changed in their meaning on account of age and experience, have convinced us that we cannot be too dogmatic in this respect. The change in the amount of correlation between similar tests from year to year is probably due to varying degrees of familiarity with the test, rather than to an age factor. The second year's inter-test correlations are slightly lower than those correlations of the first year, because, we believe, the understanding of instructions of the tests is not so important on the second year. This later factor is common to all tests and especially important at their initial presentation. It tends to raise correlations above what would otherwise be the case. After the first year there seems to be a rise in the amount of correlation between tests. We believe that the varying conditions of work in Cincinnati after the age of sixteen, and perhaps other factors, have operated to make the good slightly better, and the poor relatively poorer than they were at the age of fourteen.

VII. The practical issues drawn from these conclusions are, we believe, of great importance for future vocational testing. In the first place, we believe we have verified an opinion held by many that one well-rounded testing of an individual is likely to place his general intellectual rank for several years to come.

Secondly, we believe we have shown clearly that tested capacities in an individual may vary markedly in their stability, and that memorizing ability—if one can generalize to this extent—is a more stable function than speed or accuracy in routine work. It appears that a thorough testing out of memory will give a more permanent index of ability than other forms of testing.

School systems will, it seems probable, adopt some form of Mental Testing of their children. In order to determine what types of test are desirable and how often there should be administered, investigations along the lines indicated above will be increasingly imperative.

## APPENDIX.

### *References Cited in This Research.*

- ABELSON (1911), Br. Jr. Psy. 4, 159, "The measurement of Mental Ability of Backward Children."
- BELL, J. C. (1916), Jr. Ed. Psy. 7, 281-99, "Mental Tests and College Freshmen."
- BINET, A. (1899), *Année Psychol.* 6, 248, "Attention et Adaptation."
- BONSER, F. G. (1910), T. C. Contributions to Ed. 37. "The Reasoning Ability of Children."
- BROWN, WM. (1913), Br. Jr. Psy. 6, 223, "The Effects of 'Observational Errors' and Other Factors Upon Correlations in Psychology."
- BROWN, WM (1911), "Mental Measurement."
- BURT, CYRIL (1909), Br. Jr. Psy. 3, 94-177, "Experimental Tests of General Intelligence."
- CALFEE, M. (1913), Jr. Ed. Psy. 4, 223-31, "College Freshmen and Four Intelligence Tests."
- CHAPMAN, J. C. (1915), Col. U. Contris. to Ed. 63, "Individual Differences in Ability and Improvement, and Their Correlations."
- CAREY, N. (1915), Br. Jr. Psy. 8, 70-92, "Factors in the Mental Processes of School Children."
- HART, H. and SPEARMAN, C. (1912), Br. Jr. 5, 56-84, "General Ability, Its Existence and Nature."
- HART, H. and SPEARMAN C. (1914), Jr. Abn. Psy. 9, 217-64, "Mental Tests in Dementia."
- HOLLINGWORTH, H. L. (1913), Jr. Ed. Psy. 4, 405-14, "Correlations of Abilities as Affected by Practice."

- HOLLINGWORTH, H. L. (1914), *Psy. Rev.* 21, 1-8, "Individual Differences Before, During and After Practice."
- JONES, E. S. (1917), *Jr. Ed. Psy.* 8, 27-34, "The Woolley Tests Applied to the Detection of Ability in Telegraphy."
- KELLEY, T. L. (1915), *T. C. Contribs. to Ed.* 71, "Educational Guidance."
- KRUGER, F. and SPEARMAN, C. (1907), *Zeitsch. f. Psy.* 44, 50ff., "Die Korrelation Zwischen Verschiedenen Geistigen Leistungsfähigkeiten."
- PYLE, W. H. (1913), *Jr. Ed. Psy.* 4, 61-70, "Standards of Mental Efficiency."
- SIMPSON, B. R. (1912), *T. C. Contribs. to Ed.*, "Correlations of Mental Abilities."
- SLEIGHT (1911), *Br. Jr. Psy.* 4, 389, "Memory and Formal Training."
- SCOTT, C. A. (1913), *Jr. Ed. Psy.* 4, 509-24, "General Intelligence or School Brightness."
- STARCH, D. (1913), *Jr. Ed. Psy.* 4, 415-18, "Correlations Among Abilities in School Subjects."
- STERN, WM. (1914), "Psychological Methods of Testing Intelligence."
- THORNDIKE, E. (1914), *Educational Psychology* Vol. 3.
- THORNDIKE, E. (1915), *Am. Jr. Psy.* 27, 550 "Notes on Practice, Improvability, Etc."
- WALLIN (1916), *Psy. Rev. M. S.* 94, "Psycho-motor Norms for Diagnosis."
- WEBB, E. (1915), *Br. Jr. Psy. Mon. Sup.* 3, "Character and Intelligence."



- WELLS, F. L. (1915), *Am. Jr. Psy.* 26, 58-67, "Note On the Retention of Acquired Characteristics."
- WHIPPLE, G. M. (1914-15), "Manual of Mental and Physical Tests."
- WHITELY, M. T. (1911), *Archives of Psy.* 19, "A Study of Individual Differences."
- WINCH, W. H. 1909), *Br. Jr. Psy.* 3, 386-405, "The Transfer of Improvement in Memory in School Children."
- WOOLLEY, H. T. and FISCHER, C (1914), *Psy. Rev. M. S.* 77, "Mental and Physical Measurements of Working Children."
- WOOLLEY, H. T. (1915), *Jr. Ed. Psy.* 6, Nov., "A New Scale of Mental and Physical Measurements."
- WYATT, S. (1914), *Jr. Exp. Ped.* 2, 292-7, "The Inter-Relations of Memory Performance."
- YULE, U. (1912), "An Introduction to the Theory of Statistics."

A table to infer the probable error of Correlation for different values of indices, by the formula

$$P. E. = .6745 \frac{1 - r^2}{n}$$

(a) Where  $n$  = approximately 200—the case with all measurements in the Cincinnati series of tests, with the exception of the opposites test measurements on the first and second year.

(b) Where  $n =$  approximately 100—the case with these measurements concerned with the opposites test on the first and second years.

Value of $r$	P. E. $n=200$	P. E. $n=100$	Value of $r$	P. E. $n=200$	P. E. $n=100$
.00	.048	.067	.60	.030	.043
.10	.047	.066	.70	.024	.034
.20	.046	.065	.80	.017	.024
.30	.043	.061	.90	.009	.013
.40	.040	.057	1.00	.000	.000
.50	.036	.051			

NO. 1000  
MAY 10 1890



UNIVERSITY OF CALIFORNIA LIBRARY  
BERKELEY

THIS BOOK IS DUE ON THE LAST DATE  
STAMPED BELOW

Books not returned on time are subject to a fine of 50c per volume after the third day overdue, increasing to \$1.00 per volume after the sixth day. Books not in demand may be renewed if application is made before expiration of loan period.

JAN 13 1920

MAY 2 1920

MAY 1 1924

APR 27 1926

12 MAY '52 RL

NOV 4 1952 LO



YB 04904

376921

UNIVERSITY OF CALIFORNIA LIBRARY

